

# 《赤裸裸的统计学》

## 图书基本信息

书名：《赤裸裸的统计学》

13位ISBN编号：9787508642154

10位ISBN编号：7508642155

出版时间：2013-9

出版社：中信出版社

作者：[美]查尔斯·韦兰

页数：308

译者：曹槟

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.com)

# 《赤裸裸的统计学》

## 内容概要

# 《赤裸裸的统计学》

## 作者简介

# 《赤裸裸的统计学》

## 书籍目录

# 《赤裸裸的统计学》

## 精彩短评

1、看看

2、对于那些有数学恐惧症的人，这本书很合适。能了解统计学的基本知识，虽然很浅显，但是不枯燥。

3、相当装逼的一本书。作者活在电影里，戏真多。

4、对于文科生来讲，除了个别章节，大部分都做到了深入浅出，生动易读，基本涵盖了全部基础理论，值得推荐给所有人的统计学入门书

5、很基础的一本入门书，建议学过高数的就没必要看了。这本应该算是高中向大学的一本基础读物了

6、比较浅 有启发吧 又被图书馆扣钱了==

7、学习统计学的兴趣入门书，抛弃繁杂的公式计算，但将基本的概念讲的由浅入深，栩栩如生。喜欢。力荐。

当然，不可原谅的是，重印这么多次，书中的错误应该纠正。

这对初学者，是个伤害。

8、作者着重介绍了统计学的一些基础现象，并详细介绍了线性和回归分析这两个概念，值得对统计学不熟悉的小白阅读，例如我。

9、作者讲的例子真的超级容易明白:)

10、2016-15 对刚刚上过四节稀里糊涂统计运筹学的我来说，刚刚好。

11、这个有点入门啊

12、深入浅出，比较简单，但是例子都很好，比如那个三个门抽奖的节目

13、好书、从三门问题到正太分布等等讲的深入浅出

14、翻译的很不错，书的内容也不错。通过例子解读各种统计术语，适合高中生理解统计学术语。

163页图9-3应该是 $n = 20$ 。

164页标准误差公式应该是 $SE = s / \sqrt{n}$ 。

15、理性与感性的不完美融合~

16、作者虽然只字不提，但是始终在强调结果检验。良心读物，力荐！

17、信息时代，统计学日益重要。该书就统计学的基础原理和概念做了一番深入浅出的介绍，值得一看，也有利于提高辩证能力。

18、浅显易懂，有很多案例，在其中融合了一些统计学的思维方式，适合小白科普阅读

19、美国人写这类畅销型的普及读物，实在是厉害。

20、浅显易懂，通俗幽默，适合入门，还需再学。

21、非常好的入门读物，有些案例非常有趣。

22、通俗易懂，对于入门来说很有帮助，希望以后在统计学上多下功夫

23、挺好的有机会可以再读一遍...

24、大学统计课的案例版 统计就是不能不信也不能全信。万事万物到最后好像都是这样~

25、统计学有力量也有局限，不是所有人都会去学或从事具体的数据分析工作，重点是在数据面前应有的合理心态，就像作者说的，数学并不能代替判断。

26、很有趣也很有料的一本科普入门书，作者的逻辑，思路真是比国内很多教科书好太多了，唉。。

27、数据的陷阱无处不在，前面几章都是高中数学的内容，但是看这个比干做题有意思。译者不看NFL也可以查一下嘛，涉及橄榄球的误翻很多导致我对其他一些不了解的专有名词的翻译信任度很低。

28、通俗易懂，要是大学的时候读了此书，就不会把统计学当成数学学了。样本的准确度很重要，而我们遇到问题的时候都本能假设了样本的准确性。

29、很基础，适合没有学过的人

30、统计学的看过就忘型科普读物。

31、优秀的通俗读物。假设检验那一段看得我热泪盈眶，当年做了好多错题都没弄清公式背后的意义。不过副标题起得不好，这是讲基础统计思想的，和眼下正热的大数据无甚相关

## 《赤裸裸的统计学》

- 32、确实是好书，我以前想写的书的一部分已经被人写掉了。
- 33、很好看的统计学科普，帮我这个学渣理清了大学时稀里糊涂及格的数学课~书里面的例子还是很新的，还有奥巴马政府。。作者很幽默，要是所有教科书都这么有意思就好啦
- 34、作者贱贱的，挺好玩，有时候会义正严辞地说着一个观点，然后在下一段中马上推翻。本来想给5星的，但看别人的评论发现有些数字错误不是电子书导致的，而是实体书就错了。至于作者和中信谁背这个锅我就不管了。不值5星也值4.5了
- 35、一般吧，后面挺啰嗦的。
- 36、语言风趣幽默，感觉我又简单地把社会统计学复习了一遍。
- 37、科普用的
- 38、不清楚为啥评分这么高~~~
- 39、案例很多 能快速回忆起三年前学的的知识
- 40、8.5/10，用通俗而流畅的语言讲清楚了统计学的关键知识点、应用方法及常见误区。
- 41、不得不佩服作者的幽默和风趣
- 42、让统计学不再枯燥
- 43、挺有意思，案例多，生动有趣，就是公式少了点
- 44、非常简单的统计学，能够让人对统计学比较清楚的明白，不需要太多基础知识
- 45、好书，上学时学的统计学知识被激活。
- 46、信息泛滥年代，如何处理才是最大难题。
- 47、统计学、经济学是我们每个人都应该涉猎一些的知识，很难说上，他会让我们多富有多睿智，他们确实可以让我们改变原来固有的看待世界的角度，给我们一个全新的视角看这个世界。
- 48、感觉像看科普书一样...16%m 2周左右读完的，书中重点写中心极限定理以及其衍生出的理论，分析的十分透彻，重思想而非公式定理，包括常见的统计陷阱，调研阶段、分析维度等。一开始觉得像科普书，看着完，有了《从数据到结论》的比对才发现本书写的深刻。
- 49、额，废话有点多，给没学过概率论与统计的人看看还行
- 50、具体教我如何应用数据，详细而且为门外汉提供了实际工具，结合精益数据分析看很不错。但总觉得缺少了一部分，数据统计应该除了测试和可能性分析，还需要对比推理方面的应用吧~

1、这是一本很适合我的统计学入门书本。研究生的专业是数据挖掘，常常要用到一些统计的知识，但是由于当时急功近利的不求甚解，只求知道怎么用，不想深究为何用，以至于在这个崇尚大数据的时代，让我不能长时间地深入地全面地忽悠他人。哎，黑发不知勤学早，信口方悔读书迟！买这本书是因为专业相关，而且封面很符合我的口味。翻开一读，就像打开了一个新世界的门。这个世界首先很有趣。作者的行文风格属于生活轻幽默（突然想到的一个词）。近年来我已经很少被作家们逗笑了，不知道是见多识广还是麻木不仁，总觉得他们做作刻意。但是本书作者的文字风格确实很逗，常常让我忍俊不禁，也是让我不住读下去的动力所在。这个世界也不难懂。实例丰富、简单，而且引人入胜，通过实例能够很好的理解统计学的概念，然后又通过更好的实例讲解公式原理。这种方法让我很受用，起码我现在很清楚标准差和标准误差之间的区别，当然还有前一段很困扰我的置信区间。甚至还让我解决了我写论文时候存在的一些疑惑（当然这些疑惑没有在论文中体现）。有一种忽如一夜春风来，千树万树梨花开的豁然开朗感觉。这个世界也很广阔。我看这本书的同时，旁边放着一本统计学的专业书籍。每次看到统计学的只是概念，我感兴趣的，就去翻一翻专业书对应的章节，由于作者讲的简单易懂，勾起我的求知欲，让我意犹未尽，于是就要在专业书中把遗憾补回来，读完这本书，专业书竟然也看了不少，惊讶之余，暗暗得意，以后夸夸其谈有资本了。遇到这本书，是早就期盼而始终未得的，看完后莫名满足。众里寻他千百度，蓦然回首。

2、阅读本书，是为了更好掌握“大数据”基础。没想到，作者稍微提及下大数据的划时代好处后，就自顾自的讲起了统计学的使命。我则感觉被乞丐推销《如来神掌》，明明感觉没意义，却舍不得离开。别的不说，能把枯燥的知识讲的如此天花乱坠，作者好有一套！质量上，因为是中信出版社，天然有好感。不过书中涉及的公式因为印刷问题确实不太标准，好在都是浅知识，上学时基本都学过。而相比副标题的“数字之美”，我更相信是作者的“有趣语言”吸引我。对了，稍微介绍下作者：正常来说，你绝对猜不到他的年纪。语气上看，更是关注当下流行文化（古墓丽影等动作大片的场面描写恢弘异常），略自负（本书实际上向20世纪50年代的经典畅销书《统计数字会撒谎》致敬，作者也期望本书能经典50年）。我确信他能达到，因为作为科普入门类，本书优势实在太大了！对了，若不是作者说它1988年去了趟苏联，我还以为他刚三十而立呢~这个1966年生人的有趣的骗纸！

——谈谈武器发展史：我相信目的动机论。每个人都渴望自身价值的体现，于是开始拼搏，进步，乃至夺去，独权。当然，这是战争产生的必然。而本书中，“统计学”就是作者手中武器，天方夜谭般的案例就是诱饵。外加循循善诱的劝导。可以预见，大片读者将沦陷在有趣的统计学里不能自拔。读过本书，我确信将来会用统计学让生活更舒适真实，你们也一样~那些有趣的故事：初中第一堂政治课，老师提问了3个谜题，并宣布毕业时公布答案。实际上，我早忘记谜面了。类似的，作者在书的开篇就提问许多有趣问题。你该买彩票么？为何商城比家人更早知道你怀孕？哪些人更容易成为恐怖分子？家里电视的多少会影响学习成绩？随即，作者标注了这些答案所在的章节，没错，想知道的话只能继续阅读。而当你被通篇公式折腾的发狂时，作者又俏皮来了句“假如你没冲动的把书扔到角落，或者又走过去把书捡回来。你会发现...”没错，我继续读下去了，也确实发现...会撒谎的统计学：作为一个能让大众明白的学科，统计学对外必须是简单且有说服力的。但影响统计学的因素很多，除了比较大众的说法，还有更多“小技巧”让发布规则的人更加获益。好在书中彻底说明这种骗局是如何瞒过我们的理智勾起欲望的，不过，面对诱惑，我猜自己会再次沦陷。这才是有趣的世界。当然，互联网让统计更加便捷外，有些古老的要素也需要被证实。譬如，连续观察药物对人类的持久影响，就需要常年统计特定人群的身体状况。此外，也有许多更高深的概念等我们掌握。但从“入门级”来看，本书的吸引力搓搓有余~

3、不是认为这本书有多不好，只是远远没有心理期待的那么多。首先，不否认这本书写的通俗易懂，举了挺多生动的例子。但是，令人失望的是，作者摆弄了半天，讲的还是那么点事情，确实浅显。把这么浅显的道理讲的通俗易懂，并不是难事吧？读了整本书对统计学并没有更深的理解，最多只是稍微概念化了一点，如果去读一段统计学的定义和标准差方差等等计算，远远比这本来的更简洁。唯一收获就是这本书内关于数字会撒谎这部分内容，不同的说法基于不同的立场。（但是，这不是我读这本书的重点啊。。。）除此无他。

4、比如以州考成绩为主要依据的高中排名，排在前几位的都是一些选择性招生的学校。进入这些高中必须提出申请，其中很小一部分人（州考成绩优秀的）能入学。这一逻辑就像是给篮球队颁奖，因



为篮球队在促进队员长高方面贡献卓著。不同学生的背景、能力不同，他们的考试成绩哪些是因为自身的原因，哪些是因为学校和老师，又或兼而有之？在

5、本书适合作为高中生的科普读物。倒是深入浅出地讲解了统计学的基本特征与作用，能够勾起学生们学习统计学的兴趣。本书不适于专业人士。不过看这本书的过程中，结合自己最近的一些思考，倒是别有所得。关于书中所说的偏见与采样误差，最近倒是有一个很有趣的例子可以作为佐证。前段时间，被老婆逼着一起看了《小时代》。看完不禁大倒胃口，什么烂片，无情节，脑残逻辑，一堆所谓的帅哥靓妹在各种名牌的包装下，搔首弄姿以掩盖其空洞的大脑。后来跟朋友们吐槽，凡是看过的大多也是同样的评价。于是，我很诧异，这样的烂片，凭什么能有5亿的票房呢？直到前一阵跟一位做影视剧发行的朋友聊天，才恍然大悟。当时，我把我的疑惑向她提了出来。她给我的答案很犀利：这部片子本来就不是拍给我们这些人看的，其目标群体很明确，就是那些还在崇拜名牌、帅哥靓妹，对爱情、友情充满了憧憬的初中生、高中生、部分大学生看的。我们这些人觉得烂，不代表他们那个群体不喜欢。原来，我所询问的朋友，基本上都是与我有同样的价值观，差不多的年纪，所以大家对影片的观点也偏于一致。我的询问对象中，还真没有她所说的那些观众群体。在不知不觉中，我已经犯了选择性的采样偏差了。这就是问题所在。事实上，从投资的角度来讲，在这个市场中存在的时间够长，经验够丰富，当然是件好事。但对新事物的接受速度往往会下降。13年投资业绩最好的那批人，很多都是玩游戏的80后。也许在投资的生涯中，扩大自己的调研年龄层次也是很重要的一环。多问问新东西的吸引力到底在哪里，多尝试些新玩意儿对投资真的很有帮助。帮助自己克服各种偏见，能够更加端正地去看待很多问题。

6、稍微对本书一些在意的地方做个总结。一、数据与偏见1.选择性偏见。样本选择存在偏见。2.发表性偏见。肯定性的研究发现比否定性的研究发现更容易被发表。3.记忆性偏见。记忆会“由果推因”，没有记忆性偏见是纵向研究优于横向研究的原因之一。4.幸存者偏见。让表现差的退场。5.健康用户偏见。用户之所以健康，是因为他们本身更关注健康。二、中心极限定理中心极限定理的核心要义是：一个大型样本的正确抽样与其所代表的群体存在相似关系。虽然每个样本之间可能存在差异，单任一样本与整体之间存在巨大差异的概率是较低的。应用场景：1.如果我们掌握了某个群体的具体信息，就可以推理出从这个群体中正确抽取的随机样本的情况。2.如果我们掌握了某个正确抽取的样本的具体信息（平均数和标准差），就能对其所代表的群体做出令人惊讶的精确推理。3.如果我们掌握了某个样本数据，以及某个群体的数据，就能推理出该样本是否就是该群体的样本之一。4.如果我们已知两个样本的基本特征，就能推理出这两个样本是否来自同一群体。三、假阳性与假阴性零假设为“阴性”。1.如果检测结果推翻了零假设，而事实上并没有患病，那么即是“假阳性”，犯假阳性错误的概率即犯第一类错误（去真）的概率。2.如果检测结果没有推翻零假设，而事实上患病，那么及时“假阴性”，犯假阴性错误的概率即犯第二类错误（取伪）的概率。四、回归分析错误1.用回归方程来分析非线性关系。2.相关关系并不等同于因果关系。回归分析只能证明两个变量之间存在关系，仅凭数据无法证明其中一个变量变化就一定能导致另一个变量也发生变化。3.变量A与变量B的因果倒置。不应该使用受结果影响的解释变量。4.变量遗漏偏差。解释变量A可能会“覆盖”解释变量B，内生性。5.高度相关的解释变量。多重共线性，解释变量之间由于存在精确相关关系或高度相关关系而使模型估计失真或难以估计准确。6.脱离数据进行推断。7.数据矿。变量过多，如果取显著性水平0.05，那么每20个变量就可能出现一个假显著性。五、项目评估与“反现实”一对照组即是“反现实”1.随即控制实验。安排实验组和对照组。2.自然实验。3.非对等对照组实验。实验组和对照组之间存在一些难以察觉的差异，这些差异影响了小组成员的分配和组成，从而产生和现实有偏差的结论，这就是“非对等对照”，可以这样设置实验组和对照组：实验组为“进入名牌大学学习的同学”，对照组为“凭才华和实力可以进入名牌大学却选择去竞争没那么激烈的高校的同学”。4.差分类差分实验。首先，对某个群体接受某项介入因素或治疗之前和之后的数据进行比较，其次，将这些数据与另一个没有推出就业政策的同类组同样的指标进行比较。重要的是，用于分析的两个对象除了是否有介入因素之外，其他方面的情况基本相似，因此若两组结果之间存在显著差异，则可以认为是所评估的项目或政策的效果。5.不连续分析实验。将那些刚好符合介入或治疗条件的对象，以及以毫厘之差错失治疗机会的对象进行比较，即取两个群体的边界值。

7、昨天终于把书读完了，总的来说内容不错，翻译的也可以，毕竟是一本统计学的科普读物。觉得书的最大闪光点是把以前学的概率等统计知识与具体案例联系在一起，能够更深入的理解统计学的意义、算法和用途，能够加深对概率、相关性、极限定理、回归分析这些之前半懂不懂的知识的认识。



## 《赤裸裸的统计学》

统计学的用途就是预测未来可能发生的事情，今天成功预测了老板查岗的时间，哈哈！

8、0p 安德烈斯：用数据说谎容易，但是用数据说出真相却很难。3p 无论什么问题，统计学都极少提供唯一“正确”方法。但是，基尼系数确实以一种便捷易懂的形式为我们提供了一个重要社会现象的一些宝贵信息。8p 统计学的一个核心功能就是使用手中已有的数据进行合理推测，以回答我们还未掌握所有信息的“大”问题。简言之，我们能够使用“已知世界”的数据来对“未知世界”进行推断。8p 统计学存在的意义就是简化，因此不可避免地会丢失一些内容和细节，任何一个数字工作者对此都要心知肚明。过分依赖会带来误导性的结论或导致不良行为。11p 概率就像是武器库里的一件武器，需要使用者有较强的判断力。当然，有时候判断和赚断的界限就是那么模糊。15p 战争是为了与真实存在的敌人作战，而不是与假想敌作战(唐纳德·拉姆斯菲尔德)。学习的意义不是做题和炫耀，而是用来认清我们的生活。23p 数据越多，事实越模糊。信息过量的表格其实相当于什么都没有告诉我们。因此，我们需要简化。59p 百分率不会说谎，但它们会夸大其辞。62p 虽然“你无法管理你无法衡量的事物”，但你最好要保证你所衡量的，正是你努力想去管理的。62p 学生父母的教育程度和收入会对孩子的成绩产生不可忽视的影响。64p 提供有意义的信息与将这些信息浓缩成一个权威排名完全是两码事。在某些时候，统计学的功能仅仅是让数据看上去更顺眼。68p 利昂·波特斯坦：人们喜欢看到简单的答案。什么是最好的？当然是第一名。111p 蒙提·霍尔悖论告诉我们，对概率的本能理解有时候会将我们引入歧途。117p 金融市场的潜在风险并不像抛硬币那么容易预测，VaR这样的模型所呈现出的“伪精准”会给投资者带来虚幻的安全感。117p 失准的速度表对司机来说比没有车速表更危险：对于失准的测速表的信任或“参考”会忽略其他提示车速的不安全信息；但如果车里压根儿就没有车速表，你反而会小心地注意四周，寻找能够告诉你车辆当前行使速度的参照物。134p 绝大多数的统计学书籍都想当然地以为读者使用的都是好数据，就像每一本烹饪食谱都觉得你不会购买不新鲜的肉和腐烂的蔬菜一样。但是，如果基础数据本身就有问题，那么再缜密严谨的分析也是徒劳。136p 从直觉出发。就像从一锅汤舀出一勺进行品尝，如果之前搅拌得充分均匀，那么这小小的一勺汤足以告诉你整锅汤的味道了。144p 发表性偏见：无论在医学还是其他领域，否定性的发现都显得单调乏味。150p 如果把统计学比作侦探工作，那么数据就是线索。优质的数据就是好的线索，但首先我们必须熟悉到优质数据，而这要比看上去困难得多。218p 并不是那些“权力越大、责任也越大”的压力置你于死地，而是那种等着上司给你布置任务，但自己又没有权力决定怎么完成、何时完成这些任务的压力把人压垮了。换言之，高官们所承受的风险远远小于其秘书，因为这些助理对自己的工作任务“缺乏控制力”。218p 回归分析：在控制其他因素的前提下，对某个具体变量与特定结果之间的关系进行量化。也就是说，我们能够在保持其他变量效果不变的情况下，将某个变量的效果分离出来。239p 样本越小，结果就越分散。246p 不要用你的研究杀人。292p 每一样东西都有一个重要的功能，每一样东西都能让我们的生活变得美好，但每一样东西如果被滥用都会造成严重的后果。

9、《赤裸裸的统计学》阅读心得1. 统计学是怎么预测事情的？这件事和那件事相关，我们就可以统计这件事预测那件事。2. 当我们都无法解释原因的时候，统计学是怎么找到真正原因的？通过统计的数据规律发现结果，从结果中倒回来找原因。知道结果找原因要比什么都不知道找原因要快得多，因为我们天生喜欢给结果找原因，有动力又有方向。3. 平均数是怎么骗我们的？因为受到了强大“异常值”的影响，强烈拉高或拉低了平均值。解决方法是同时看中位数和平均数，“中间位置”往往具有“集中趋势”。二者差距异常必然有问题。4. 百分数（相对数）是怎么骗我们的？百分数必须考虑基数大不大，基数太小会显得提升/降低的百分数很惊人，基数太大会显得提升/降低的百分数很平常。我们要看的不是增长了多惊人的百分比，而是具体从什么数增长到了什么数。只拿出惊人的或平常的百分数又不告诉你基数的，也许在忽悠你。5. 中位数是怎么骗我们的？与平均数相比，中位数没有考虑强大“异常值”的影响，也许你就是偏向强大“异常值”呢，这种情况中位数就反映不出你的信息。解决方法是同时看中位数和平均数，二者差距异常则想想强大“异常值”。有人只拿出中位数或平均数一个给你看有可能是别有用心。6. 与金钱有关的数据（票房数据增长、福利数据增长和最低工资数据增长）是怎么骗我们的？与钱有关的都要考虑一个重要问题——通货膨胀的影响。考虑了通货膨胀调整的实际数据和没有调整的名义数据差距非常大。给你的钱多了别高兴，也许购买力还下降了。7. 机构排名是怎么骗我们的？有些统计项目没有意义（比如优秀学生统计，有些学校只招优秀学生）或者有意义却难以统计（比如高校排名的“学术荣誉”），导致机构排名不合理，可能排名靠后的机构反而让你以较小成本获得较大收益。8. “盲品测试”为什么风险不大？所谓“盲品测试”是摆出相似产品（比如两种品牌啤酒），让某一品牌忠实客户在事先不知道的情况下测试自己喜欢哪一种。

## 《赤裸裸的统计学》

这种测试风险不大，因为大部分客户其实很难分辨两种产品的细微差别，结果就是抛硬币50%的概率。而商家只要表示另一品牌忠实客户有近50%喜欢自己的就达到营销目的了。9. 为什么赌场永远是最终的赢家？博彩业的每一种玩法规则定下来，概率就定了下来，定下的概率对赌场有利，长远来看赌场就是最终的赢家。足够长的时间足够多的次数，概率总是最终的赢家。（同样，保险公司也会是最终的赢家）10. 有人找我们玩“游戏”，该不该冒险？面对不确定，看期望值。把每一种可能与其对应的概率相乘并把所有结果相加汇总，得出玩这个“游戏”的期望值。付出的成本小于这个期望值就玩，大于就不玩。次数越多，结果越趋向于期望值。（推荐学习一下决策树形图）11. 统计学本身存在什么偏差？不负责任的数据使用：某些机构的“伪精准”数据、不知道事件之间的联系：有联系的事件连续发生的概率比独立事件大得多、选择性偏见：调查的对象拥有强烈的主观意愿、发表性偏见：给人们看的都是好的、记忆性偏见：回忆得到的是印象深刻的，不一定是真实的、幸存者偏见：走了一部分低端值，幸存的高端值拉高了平均值。还有诸多偏见等等。

10、通过生动的故事和生活化的语言，介绍统计学的几种基本方法，让人发现用统计学的观念去看待社会生活中的现象原来是这么有意思，潜移默化地告诉我们，大数据时代，各种花花绿绿的数据和学说是否欺骗了我们？或是否作者自己也拿不准就在那里信誓旦旦？

## 章节试读

### 1、《赤裸裸的统计学》的笔记-第6页

(第XI页) 瑞典数学家、作家安德烈斯的一句话：用数据说谎容易，但是用数据说出真相却很难。

### 2、《赤裸裸的统计学》的笔记-第143页

发表性偏见。肯定性的研究发现相比否定性的研究发现来说，更有可能被发表，从而影响我们对事实真相的判断。假设你刚刚完成了一项严谨的纵向研究，得出玩电子游戏不能预防结肠癌的结论...

...  
没有一本杂志会发表你的研究成果，原因有二。首先，玩电子游戏和患直肠癌(p.144)之间在科学上并不存在强烈的相关性，因此你开展这项研究的出发点是什么并不明确。其次，也是本节重点要讲的，一个为“某因素不能预防癌症”的事实算不上是一个特别有趣的发现，毕竟有太多的因素都不能预防癌症。.....

而这种发表性偏见将会导致研究结果的扭曲。假设你的研究生班的一个同学开始了一项不同的纵向研究，她发现每天花很多时间玩电子游戏的人患上直肠癌的概率确实低，这样的结论就有趣多了！只有这类发现才能吸引眼球，容易受到医学杂志、大众媒体、博客以及电子游戏厂商.....的关注。对于统计学来说，巧合的存在决定了异常事情的发生.....在100项完成的研究中，就有可能存在一项结论完全不值得信赖的情况，比如玩电游可以降低患直肠癌的风险这样的结论。但.....99项证明电游和直肠癌之间不存在任何联系的研究成果由于枯燥无趣，将不会得到发表；但剩下的那一项声称寻找到二者之间联系的论文会引起注意并最终得到发表，仅仅因为这样的结论是有趣的。偏见的来源并不是研究本身，而是那些能够接触到大众读者的信息载体，于是阅读科学杂志的人会读到这篇关于电子游戏和癌症的文章，在他们的眼里关于这个话题只有一项研究，而且这项研究证实玩电子游戏的确和癌症有关，事实上，有99%的研究证明二者并不存在联系。

(p.145) 为了解决这一问题，如今的医学杂志要求所有研究在刚开始时通过项目注册的方式予以告知，否则将取消其出版的资格，杂志编辑可以借此得出某项研究的肯定和否定结论的比例。

### 3、《赤裸裸的统计学》的笔记-综合笔记

#### 数据说谎之偏见

#### 1、选择性偏见

某年，总统选举时民意调查机构通过电话调查得到数据，殊不知该年拥有家庭电话的都是富裕家庭，数据自然要造成偏差。

#### 2、发表性偏见

为数据发表而篡改数据。

如今医学杂志要求所有研究在刚开始时通过注册项目予以告知，借此可以得到肯定和否定的比例，用以判断真假。

#### 3、记忆性偏见

乳腺癌患者在回忆饮食习惯时，会倾向于夸大高脂肪饮食的摄入。她们不断回忆过去，想从中寻找到患病原因，然后再将原因植入记忆。

## 4、幸存者偏见

成绩差的学生会逐渐辍学，导致学校内平均成绩上升，但其实教学质量并未有任何改善。

## 5、健康人偏见

会定期服用维生素的人更健康，因为本来会服用维生素的人就是更关心和在意健康的人。

分离样本测试：将同一个问题的不同问法用在不同的样本人群里，以便判断用词上的小变化会否影响回答者的答案。

为避免受访者言不由衷，可以采用迂回的提问方式，如：你身边有认识的人持有这种观点吗？

拥有警察最多的地方通常是犯罪率最高的地方，但这并不代表是警察导致犯罪率，或警察对犯罪药石无效。

## 4、《赤裸裸的统计学》的笔记-第45页

马克·吐温有一句名言是这样说的，“谎言有三种：谎言、该死的谎言，以及统计学”。

## 5、《赤裸裸的统计学》的笔记-第146页

用回归方程式来分析非线性关系  
相关关系并不等同于因果关系  
因果倒置  
变量遗漏偏差[遗漏了重要变量]  
高度相关的解释变量（多元共线性）  
脱离数据进行推断[超出数据适用范围]  
数据矿（变量过多）

## 6、《赤裸裸的统计学》的笔记-第52页

数据本身并没有撒谎，只是有些数据没有发出声音罢了。

## 7、《赤裸裸的统计学》的笔记-第159页

4. 中心极限定理告诉我们，……样本平均值会在群体平均值周围呈现一个正态分布……有68%【68.2%】的数值位于平均值一个标准差的范围之内，95%【95.4%】的数值在平均值两个标准差范围内，【99.7%的数值在平均值三个标准差范围内】以此类推。

5. 不论所研究的群体是怎样分布的，上述结论始终都是成立的。就算样本所在的群体不是正态分布，也不影响其样本平均值的正态分布形态。

……  
(p.167) 要引起注意的是，为了能够让中心极限定理成立，样本数量必须足够多（依照经验法则，至少有30个）。

……  
(p.168) 如果出现了某个概率较低的结果，我们便可以推测是不是有一些其他因素介入，而且概率越低，其他因素介入的可能性就越大。



# 《赤裸裸的统计学》

这些基本上囊括了统计推断的所有内容，而中心极限定理是让这一切发生的重要推动力。

## 8、《赤裸裸的统计学》的笔记-第46页

## 9、《赤裸裸的统计学》的笔记-第188页

【I型错误=假阳性】如果我们用于推翻零假设的举证责任定得过于宽松（例如0.1），那么我们就（p.189）会经常处于推翻零假设的状态。而实际上，在很多时候零假设是正确的。这就是统计分析中肯定或否定假设前提的I型错误。……I型错误表示错误地推翻了一个零假设……我们也称之为“假阳性”。……

【II型错误=假阴性】……推翻零假设的门槛越高，我们推翻零假设的可能性就越小，以至于很多应该被推翻的零假设“逃过一劫”。……（p.190）这就是统计学上的II型错误，又称为“假阴性”。哪种错误更加严重？这要依情况而定。最重要的是，你能够意识到宽松和严格之间的权衡妥协。

## 10、《赤裸裸的统计学》的笔记-第109页

（p.107）“蒙提 霍尔悖论”是一个著名的概率难题。1963年美国开播的电视游戏节目《让我们做个交易》……每一期节目播到最后，总会有一个参赛者脱颖而出，站在主持人蒙提 霍尔旁边，在他们的眼前有3扇巨大的门，编号分别为1、2、3。……

……2008年，《纽约时报》专栏作家约翰 泰拿尼专门就“蒙提 霍尔现象”写了一篇文章。随后这份报纸还在网站上开辟了一个互动专题，读者可以亲身体验这个游戏，包括提示你是否要改变选择，游戏的最后甚至还有可爱的小羊和小轿车从门后跳出来揭晓答案。这个游戏会记录下你改变和坚持最初选择的成功率，你可以试一下。

【网站互动专题】

## 11、《赤裸裸的统计学》的笔记-第232页

[（p.161）密歇根大学主持了一项名为“变化的一生”的纵向研究，对几千名美国成人的各项指标进行了监测，其中就包括他们的体重。]

教育与体重呈现负相关关系。在“变化的一生”项目的所有研究对象中，受教育时间每增加一年，体重就相应减少1.3磅。

接受政府食物补助的个人要比其他人重。在其他因素保持不变的条件下，接受补助的人要比其他研究对象平均重5.6磅。

种族变量是其中最有趣的变量。就算将上述所有因素[身高、年龄、性别、运动等]都控制起来，种族因素依然对体重有着举足轻重的影响。“变化的一生”参与者中非西班牙裔成年黑人要比其他人平均重10磅。

## 12、《赤裸裸的统计学》的笔记-第228页

当回归系数至少是标准误差的两倍或以上的时候，该系数极有可能具有统计学意义。

## 13、《赤裸裸的统计学》的笔记-第93页

## 14、《赤裸裸的统计学》的笔记-第15页

## 《赤裸裸的统计学》

.....这暴露了回归分析的一个局限所在。我们可以通过统计分析来确定两个变量之间的强烈联系，却无法解释为什么存在着这样的联系，在某些情况下，我们也无法确定这种关系是否为因果关系...。在恐怖主义的例子中，克鲁格教授[在《恐怖分子从何而来？》书中]推测，由于恐怖分子的行动一般都带有政治目的，所以只有受过高等教育和家庭殷实的人才会有最大的动力去改变社会，这些人尤其忍受不了某些政府部门对自由的压制，从而走向恐怖主义。根据克鲁格教授的研究，在其他因素相同的前提下，恐怖活动频繁出现的国家往往是那些实行高压政策的国家。

### 15、《赤裸裸的统计学》的笔记-第145页

记忆性偏见。回忆确实很神奇，但并不是优质数据的可靠来源。我们总是认为现在和过去是有逻辑联系的.....当我们试图解释当前一些特别好或特别坏的结果时，我们的记忆便会出现“系统脆弱”的尴尬。1993年，一位哈佛大学的研究人员进行了一项关于饮食习惯和癌症关系的研究.....（p.146）《纽约时报》是如此形容这一记忆偏见的“阴险本质”的：  
一纸乳腺癌的诊断书不仅改变了一个女性的现在和未来，还改变了她的过去。.....不断地回忆过去想要从中找到一个患病原因，然后再将这个原因植入记忆。

没有记忆性偏见是纵向研究优于横向研究的一个方面。

### 16、《赤裸裸的统计学》的笔记-第76页

我们必须牢记一点，那就是相关关系并不等于因果关系。

### 17、《赤裸裸的统计学》的笔记-第38页

10、25、50、75、90分位数来描述有偏分布

### 18、《赤裸裸的统计学》的笔记-第161页

出色的研究者.....能够找到有创意的方式来控制变量，使得对照实验能够进行下去。在对某项疗法或某种介入手段进行测量时，我们需要制定一些类似于标尺的东西来提供参照。.....聪明的研究者就能找到有创意的方式来测量某种疗法或介入手段的影响，也就是在现实与“反现实”（在介入手段缺失的前提下所发生的结果）之间搭建比较的桥梁。

#### 【方法】

随机控制实验[警觉试验]

自然实验

非对等对照实验

差分类差分实验[首先对某个群体接受某项介入因素或治疗之前或之后的数据进行比较，其次将这些数据与另一个没有接受该介入因素的群体的同期数据进行比较]

不连续分析实验[将那些刚好符合介入或治疗条件的对象，以及以毫厘之差错失治疗机会的对象进行比较]

### 19、《赤裸裸的统计学》的笔记-第52页

（对于小布什政府对其减税政策的说法）《纽约时报》评价说：“数据本身并没有撒谎，只不过有些数据没有发出声音罢了。”

### 20、《赤裸裸的统计学》的笔记-第97页

你应该时刻谨记为那些你无法轻松承受的意外上保险，而其他情况就不要浪费钱了，这是个人理



财的核心原则之一。

## 21、《赤裸裸的统计学》的笔记-引言

我天生就很排斥数学。我对数字本身没有任何好感，对那些在现实世界中毫无用处的骗人公式也没有什么好印象。我尤其不喜欢高中的微积分课，原因很简单，因为从来就没有人告诉过我学习这门课的意义是什么——有谁会在乎抛物线下方的区域代表什么？

。。。有趣的是，尽管物理课也需要进行像微积分课那样令人厌烦的演算，但我在高中时却十分喜欢物理课。这又是为什么？因为物理课有一个明确的目的。。。

上大学之后，我彻底沉醉于概率学中，因为它同样为我在洞察现实生活中的一些有趣场景提供了解释。回想过往，我意识到让我痛恨微积分课的不是数学，而是从来就没有人想到要告诉我数学的意义是什么。如果你没有被“高雅”的公式本身所吸引——反正我是一点儿都不觉得有什么“高雅”的——那么，你面对的只会是繁冗而机械的公式，至少我的老师当初就是这样把它们教给我的。也正是因为这一点，我与统计学结了缘。我爱统计学。

# 《赤裸裸的统计学》

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu111.com](http://www.tushu111.com)